



JVET-J0019

Description of 360° video coding technology proposal by MediaTek

Jian-Liang Lin, Ya-Hsuan Lee, Cheng-Hsuan Shih, Sheng-Yen Lin,
Hung-Chih Lin, Shen-Kai Chang, Peng Wang, Lin Liu, and Chi-Cheng



Overview Summary

- **Modified Cubemap Projection (MCP)**
 - Compact 3x2 layout without padding
 - Face size: 1184x1184 (1280x1280 is also tested)
- 360° specific coding tools
 - Inter prediction
 - Intra prediction
 - In-loop filters
- Experimental Results:

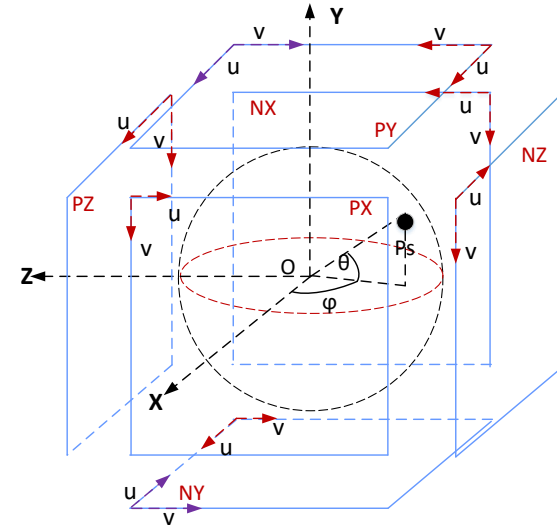
Face size	Coded sample ratio	Anchor	E2E WS-PSNR			
			Y	U	V	YUV
1184x1184	99.88%	HM	-35.5%	-69.5%	-71.6%	-44.3%
		JEM	-15.8%	-50.7%	-52.6%	-24.8%
1280x1280	116.73%	HM	-36.2%	-69.4%	-71.7%	-44.8%
		JEM	-16.8%	-51.3%	-53.4%	-25.7%

Outline

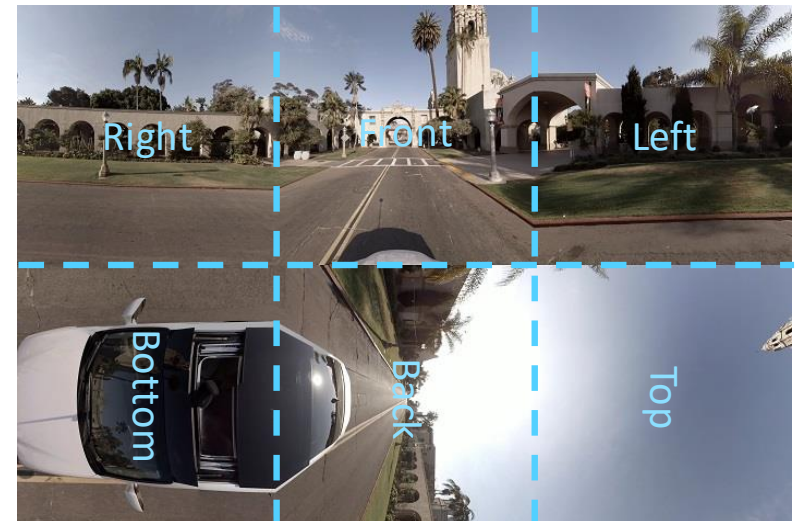
- Modified Cubemap Projection (MCP)
- 360° specific coding tools
 - CU Partition
 - Inter Prediction
 - Geometry Padding
 - Motion Vector Projection
 - Intra Prediction
 - Neighboring CU Availability
 - 360°-based In-loop Filters
- Compression Performance
- Complexity Analysis

Cubemap Projection

- Definition of the cubemap projection is the same as that in 360Lib.
- Each face is defined by a two-dimensional (u, v) plane with u and v being in the range of $[-1, 1]$.

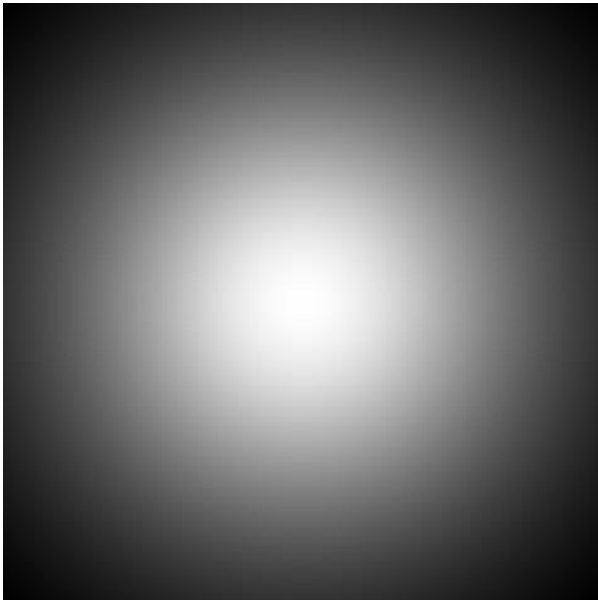


Face index	Face label	Description
0	PX	Front face with positive X axis value
1	NX	Back face with negative X axis value
2	PY	Top face with positive Y axis value
3	NY	Bottom face with negative Y axis value
4	PZ	Right face with positive Z axis value
5	NZ	Left face with negative Z axis value



Cubemap Projection

- Drawbacks of the conventional CMP
 - Inefficient sampling rate inside a face
 - Turning points appeared when a straight line crosses face boundaries



WS-PSNR weighting

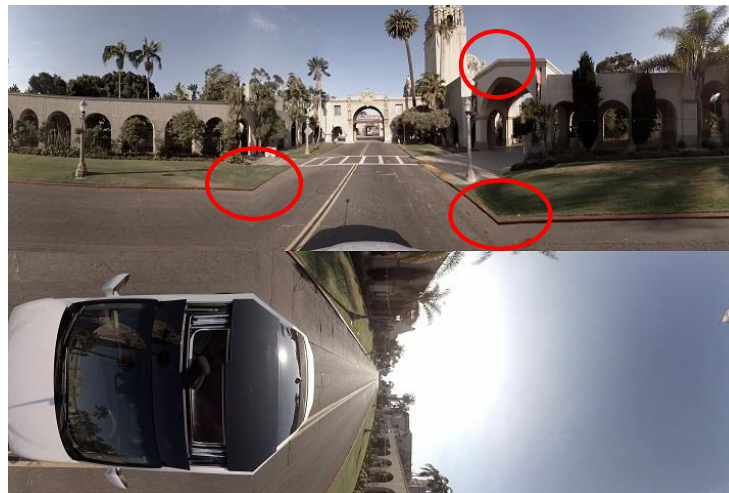


Equal-Angular Cubemap

- The equal-angular cube (EAC) projection applies an equal-angular mapping function to achieve uniform sampling distribution.

Face label		EAC
2D-to-3D	u -axis and v -axis	$f(x) = \tan\left(\frac{\pi}{4} x\right)$
3D-to-2D	u -axis and v -axis	$g(x) = \frac{4}{\pi} \tan^{-1}(x)$

- The uniform sampling distribution is applied on each axis separately, not jointly.
- The turning points still appeared on the face boundaries.

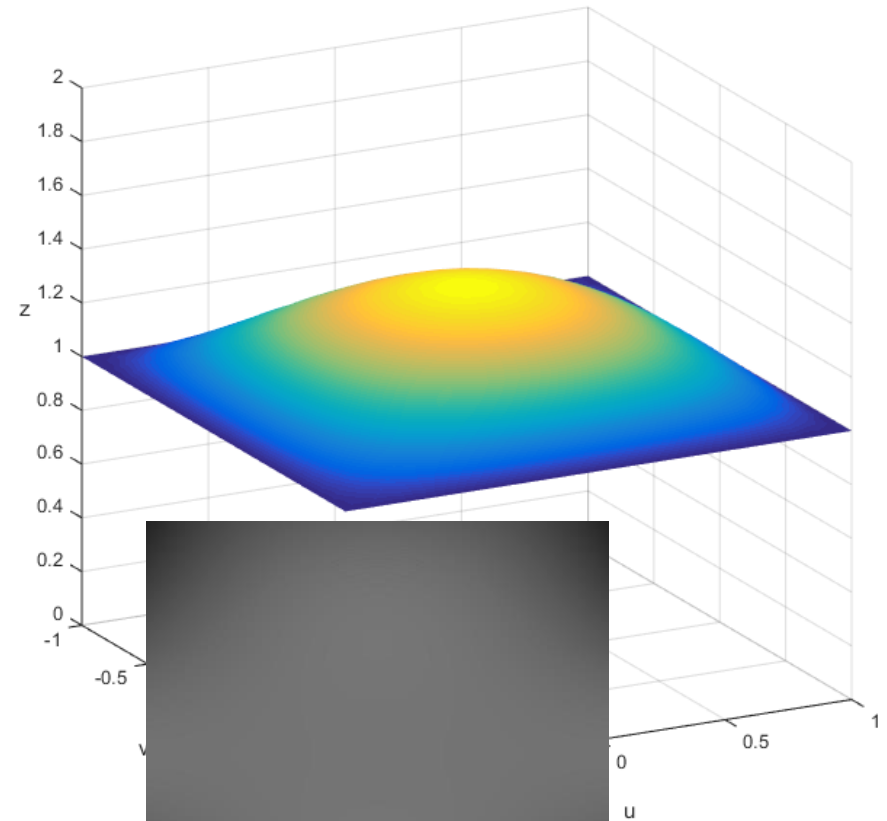


Modified Cubemap Projection

- Introduces a coordinate mapping function
 - $z = 1 + 0.4(1 - u^2)(1 - v^2)$
- Radially decrease sampling rate from face center to face corner



WS-PSNR weighting of CMP



WS-PSNR weighting of MCP

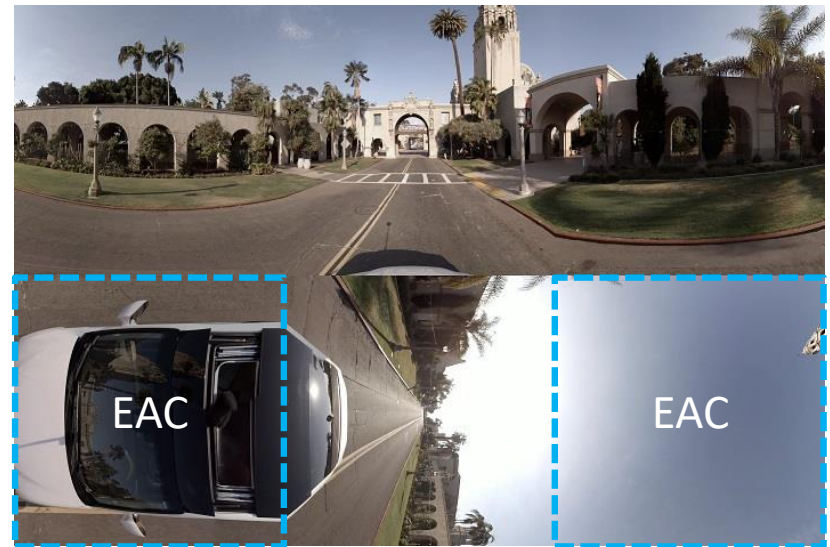
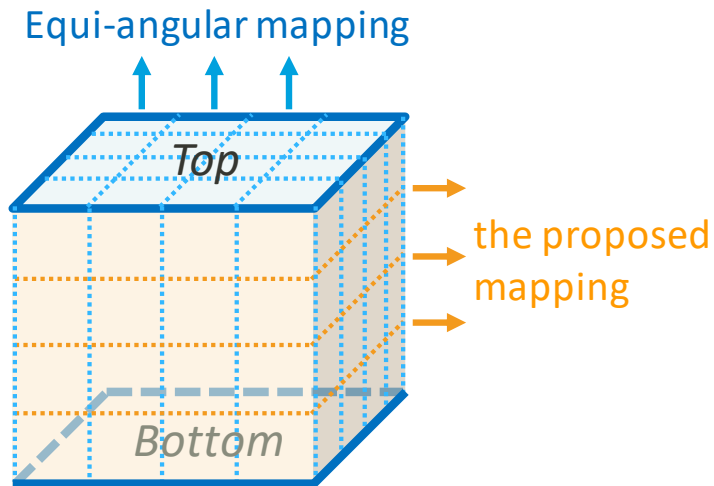
Modified Cubemap Projection (MCP)

- For the Front, Back, Right and Left faces, the proposed mapping function is applied on v -axis (corresponding to the latitude) to replace the equal-angular mapping function in EAC.

Face label		EAC & Top/Bottom Faces in MCP	Front, Back, Right and Left Faces in MCP
2D-to-3D	u -axis	$f(u) = \tan(\frac{\pi}{4}u)$	$f(u) = \tan(\frac{\pi}{4}u)$
	v -axis	$f(v) = \tan(\frac{\pi}{4}v)$	$f(v) = \frac{v}{1 + 0.4(1 - u^2)(1 - v^2)}$
3D-to-2D	u -axis	$g(u) = \frac{4}{\pi} \tan^{-1}(u)$	$g(u) = \frac{4}{\pi} \tan^{-1}(u)$
	v -axis	$g(v) = \frac{4}{\pi} \tan^{-1}(v)$	$g(v) = \begin{cases} v, & \text{if } t = 0 \\ \frac{1 - \sqrt{1 - 4t(v-t)}}{2t}, & \text{otherwise} \end{cases}$ where $t = 0.4v(g(u)^2 - 1)$

Modified Cubemap Projection (MCP)

- The Top and Bottom faces apply the equal-angular mapping on both u -axis and v -axis to keep the continuity with other faces.



360° Video Specific Coding Tools

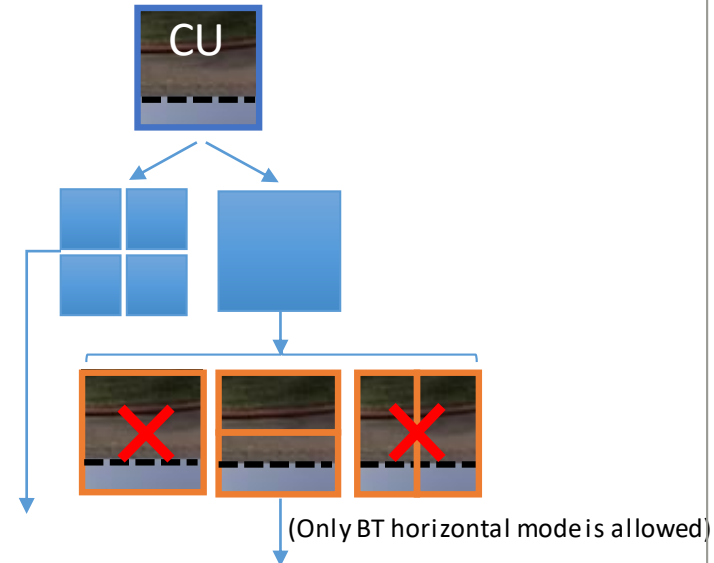
- Because of the **geometric continuity** for 360° video, the following coding tools are appropriately changed to improve the coding efficiency.
 - CU partition
 - Inter Prediction
 - Geometry Padding
 - Motion Vector
 - Intra Prediction
 - Neighboring CU Availability
 - 360°-based In-loop Filters

CU Partition

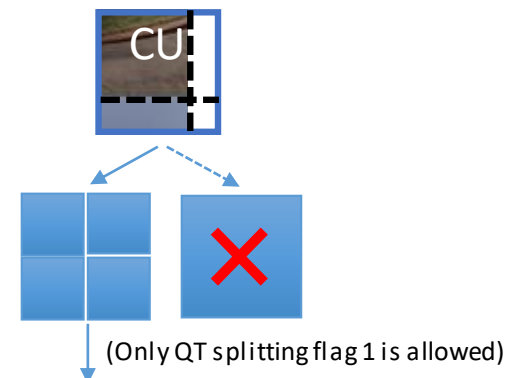
- To avoid a CU containing the discontinuous edge
- Automatic CU split for discontinuous edge is proposed.



Case 1 CU with horizontal discontinuous edge inside

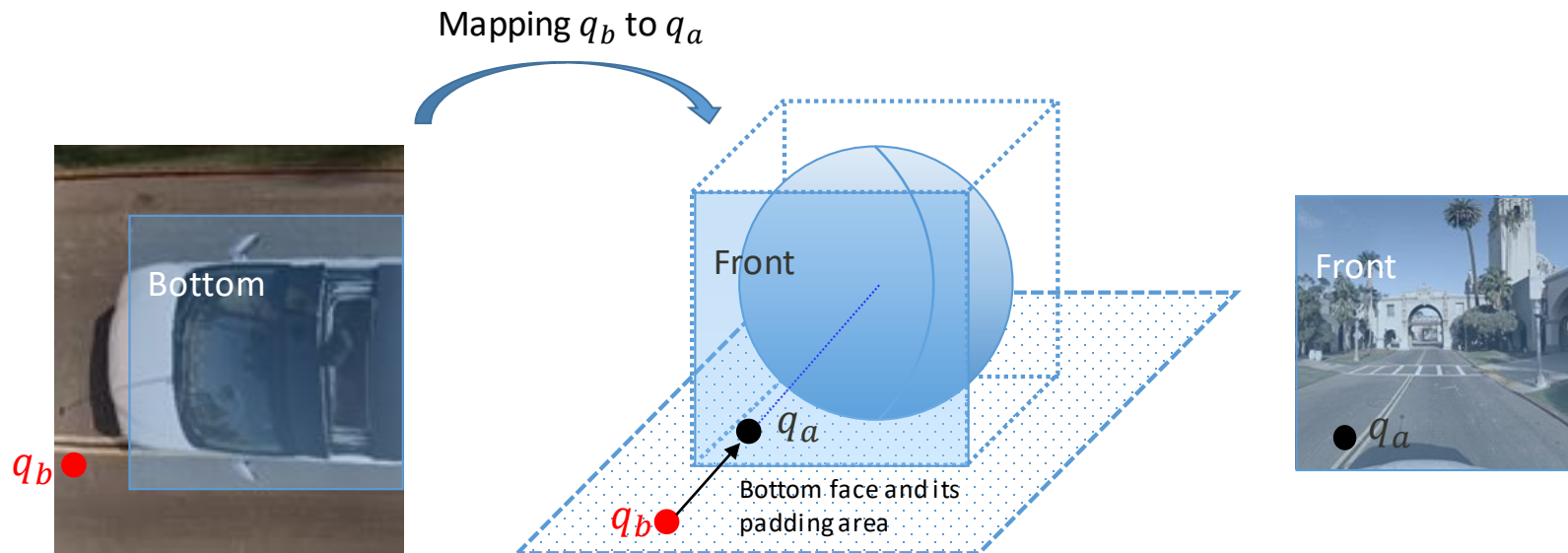


Case 2 CU with horizontal discontinuous edge and vertical frame boundary inside



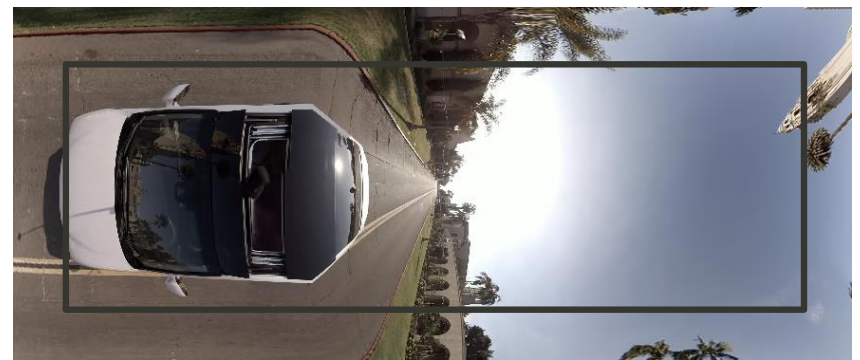
Geometry Padding

- **Geometry padding** is used as the reference pixels for the inter prediction when the reference pixels are outside the frame boundary or across the discontinuous edge.
- When the derived coordinate q_a is fractional, the 6-tap and 4-tap Lanczos filters are applied for luma and chroma, respectively.



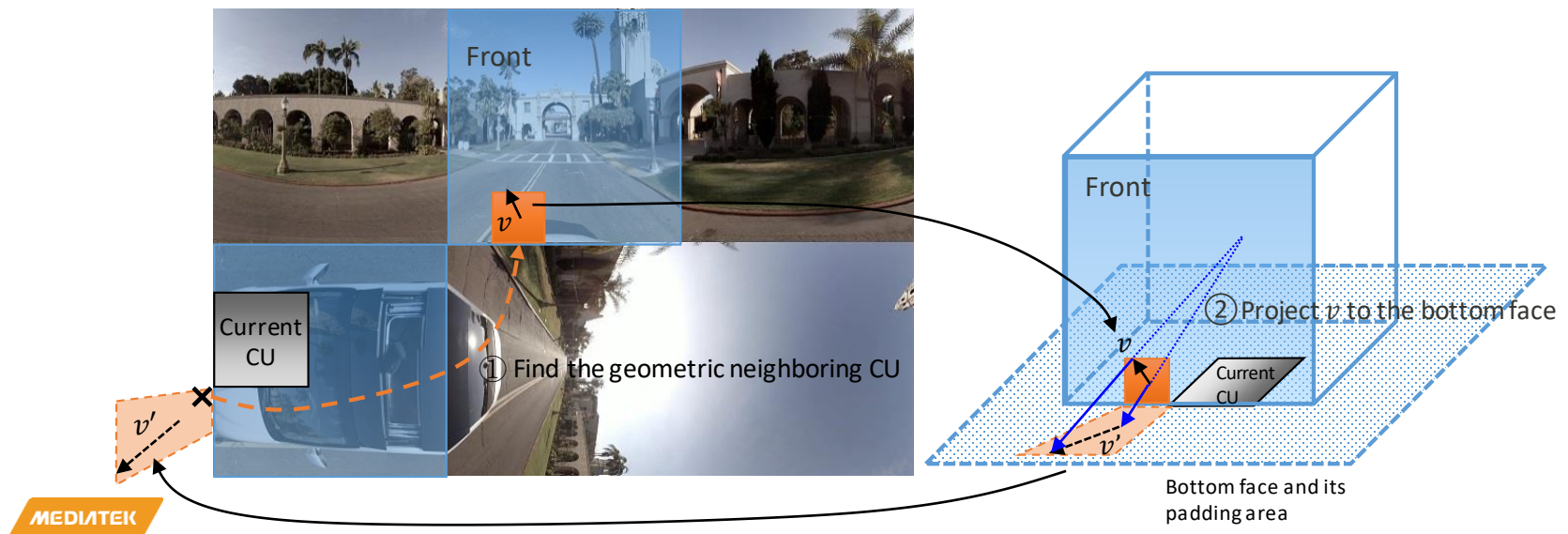
Frame Buffer for Geometry Padding

- The proposed implementation requires two sub-frame buffers.
- Each sub-frame buffer has one 3x1 face region with an extended padding area.
 - Padding width is one LCU width with additional 16 pixels: 272 pixels



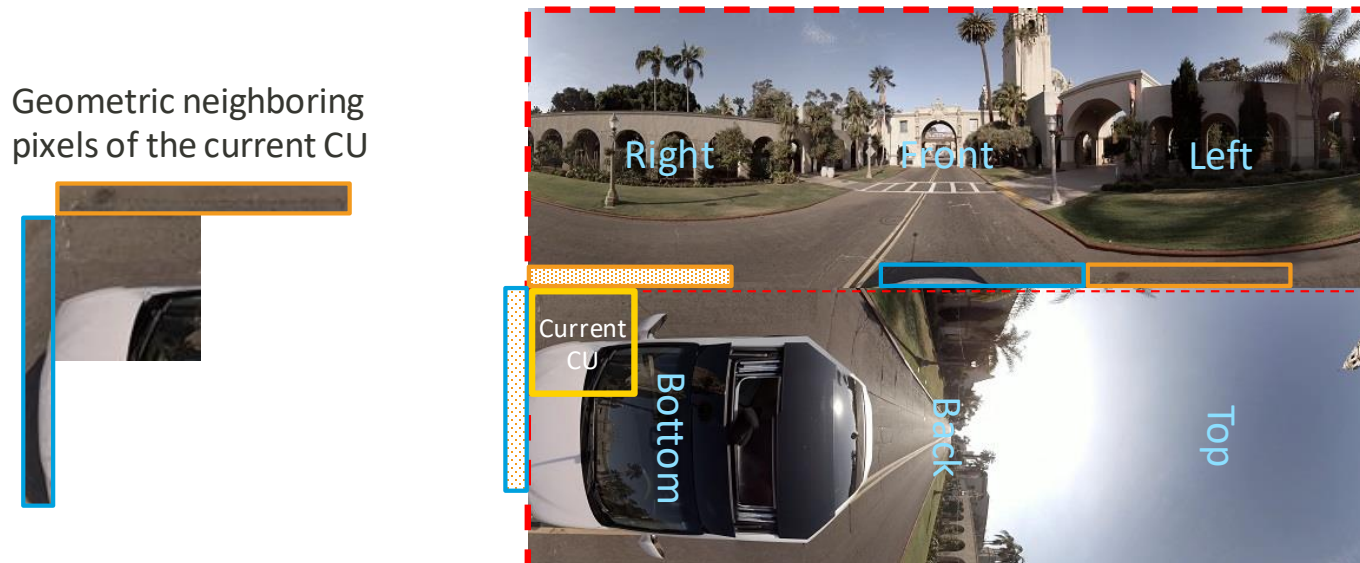
Motion Vector Projection

- When a CU is located beside the frame boundary or the discontinuous edge, its neighboring CU is replaced by the **geometric neighboring CU** located at other faces.
 - Determine a geometric neighboring CU according to the layout in MCP
 - Project the motion vector of the geometric neighboring CU onto the extended face at which the current CU is located



Intra Prediction

- The intra prediction constructs efficient predictors by using the neighboring pixels of the current CU.
- The neighboring pixels are replaced by the **geometric neighboring pixels**, which are located at other faces, when a CU is located beside the frame boundary or the discontinuous edge.
- Note that the geometric neighboring reference pixels are unavailable when these pixels have not been coded yet.



Neighboring CU Availability

- When the neighboring CUs are located outside the frame boundary or across the discontinuous edge, those neighboring CUs are considered as unavailable to avoid inappropriate reference.
- The neighboring CU's availability check is applied to all the other tools except the motion vector prediction.
- For motion vector prediction, the geometric neighboring CU derivation is applied.

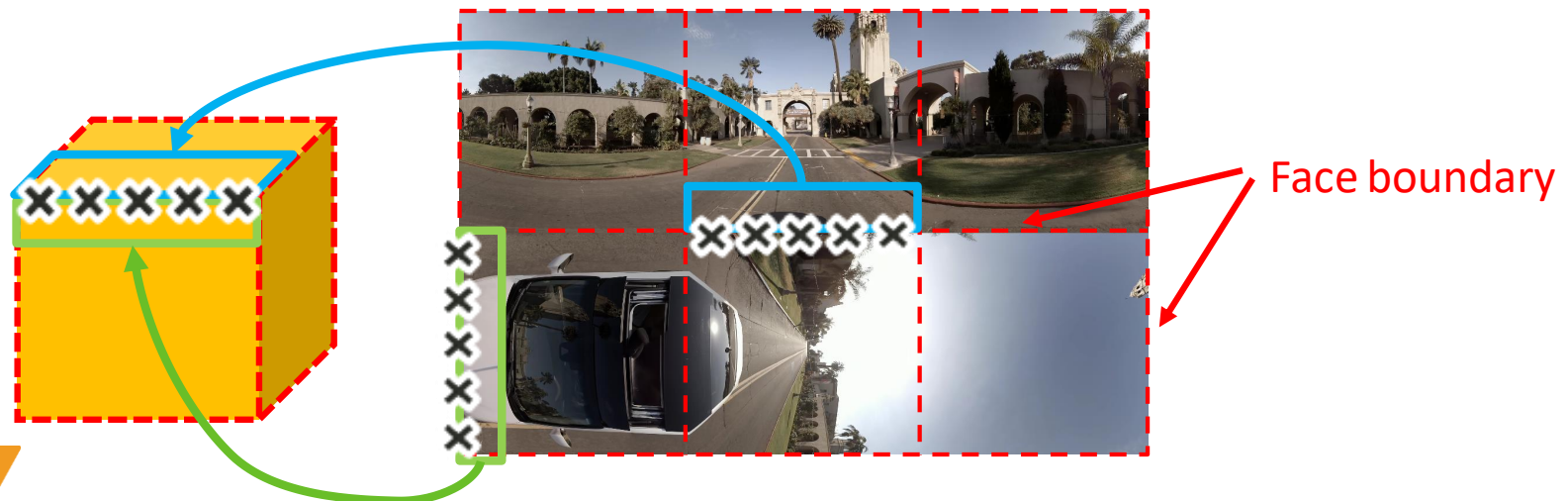


360°-based In-loop Filtering

- Because of the **geometric continuity**, the in-loop filtering processes should be modified accordingly.
- Use the pixels data located at the geometric positions for filtering, when the accessed pixels are outside the frame boundary or across the discontinuous edge.
- The modified filters include:
 - De-blocking Filter
 - Sample Adaptive Offset Filter
 - Adaptive Loop Filter
 - Non-Local Means Filter (proposed in J0018)
 - Convolutional Neural Network Filter (proposed in J0018)

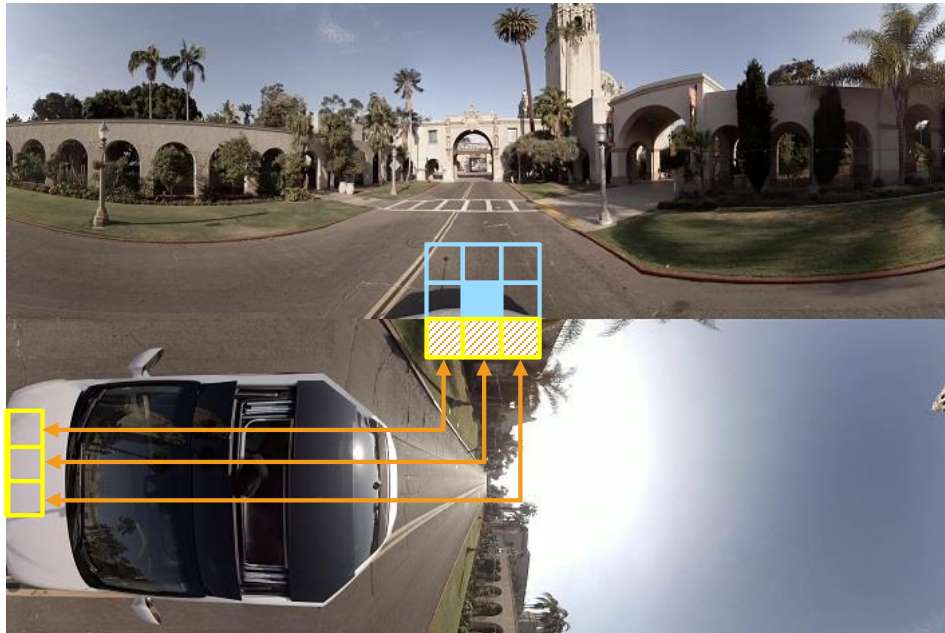
360°-based De-blocking Filter

- Fetch the **geometric pixel** located at another face for de-blocking when the filtered pixels cross the frame boundary or the discontinuous edge.
- The **pixel data** in the geometric positions are copied directly from the corresponding face.
- The **motion vector** located at the geometric pixel positions is also adjusted by the MV projection for appropriate boundary strength calculation.



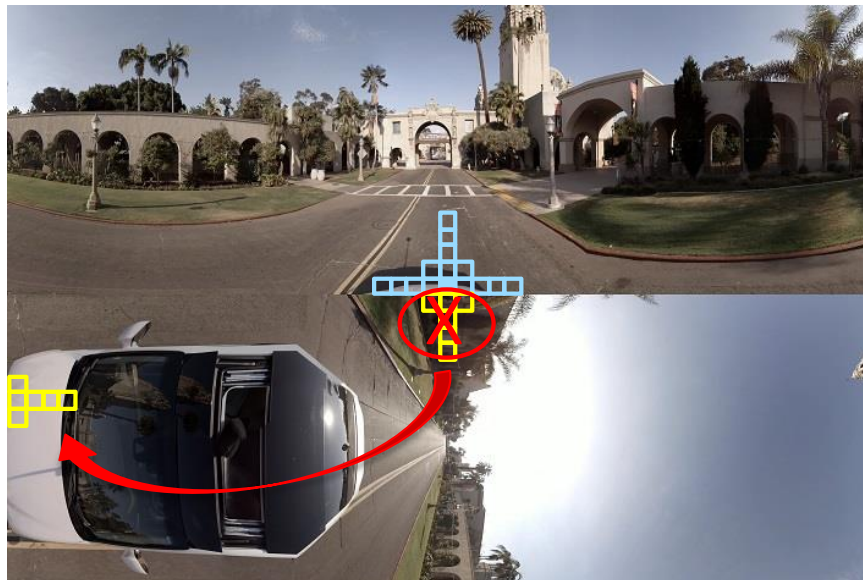
360°-based Sample Adaptive Offset Filter

- When the referenced neighboring pixels cross the discontinuous edge
- Fetch the **geometric pixel** located at another face for SAO filtering
- The pixel data in the geometric positions are copied directly from the corresponding face.



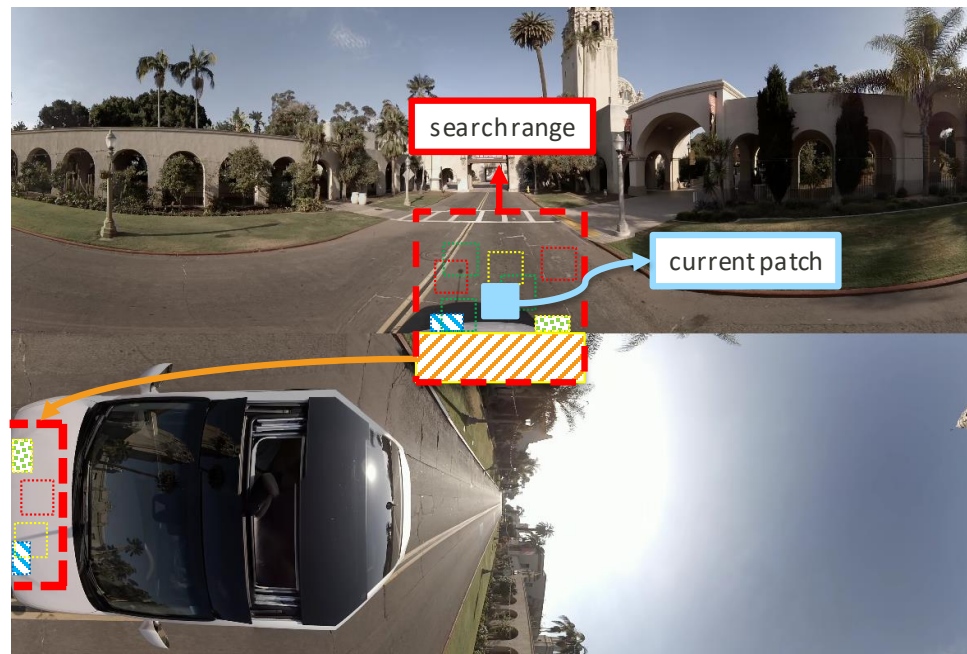
360°-based Adaptive Loop Filter

- In the conventional ALF, the replication padding is used when the accessed pixels for filtering are located outside the frame boundary.
- When the referenced pixels are across the frame boundary or the discontinuous edge
- The proposed modification replaces the referenced pixels with the **geometric padding** pixels derived from other face.



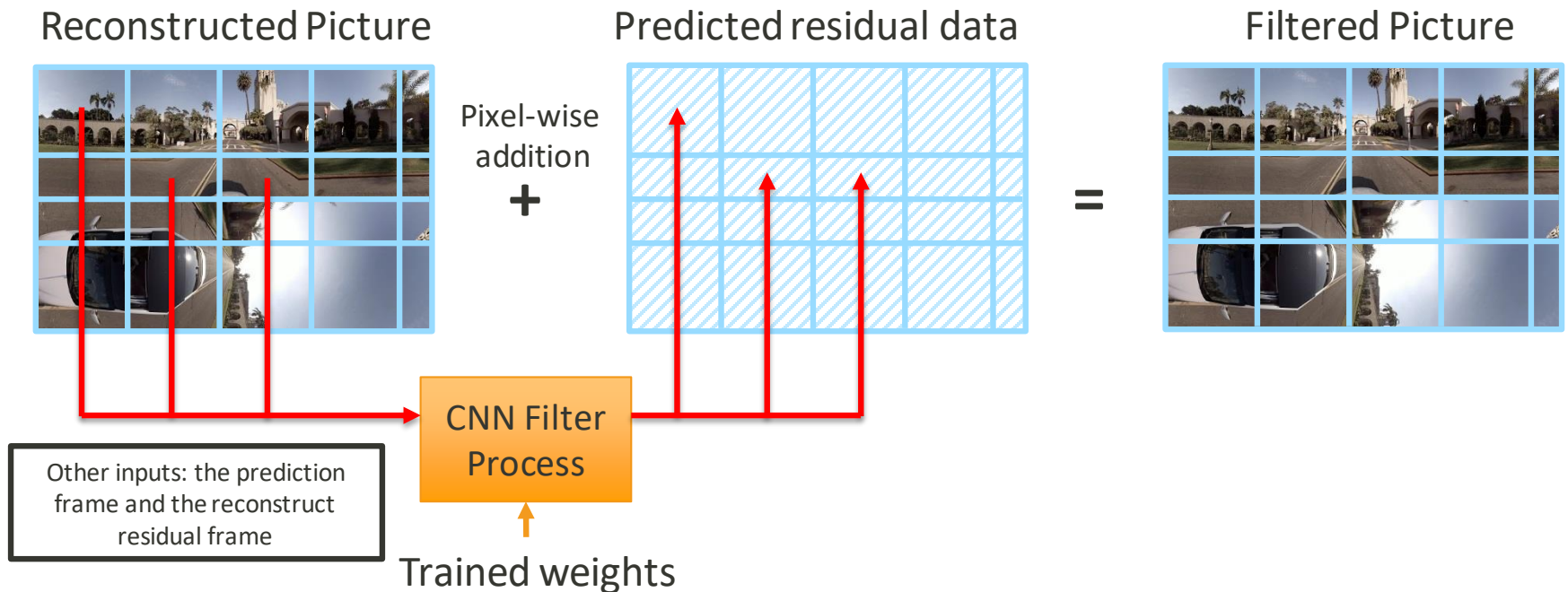
360°-based Non-Local Means Filter

- The NLM collects multiple similar group of pixels (patches) to reduce the coding error.
- Search multiple similar patches within a **geometry continuous region**
 - search range: $[-16, 16]$
 - patch size: 8×8



360°-based Convolutional Neural Network Filter

- For the CNN filter, a processing unit is divided into two small processing units in the filtering process when containing the discontinuous edge.
- Separately processed by the CNN filter to avoid referencing the spatial-irrelevant data.



Subjective Quality

Balboa (frame 29) with Viewport Center(135, 35)



Directly use neighboring pixels for filtering



Use geometric neighboring pixels for filtering

Subjective Quality

ChairliftRide (frame 211) with Viewport Center(-45, -35)



Directly use neighboring pixels for
filtering



Use geometric neighboring pixels for
filtering

Subjective Quality

Trolley (frame 171) with Viewport Center(135, 35)



Directly use neighboring pixels for filtering

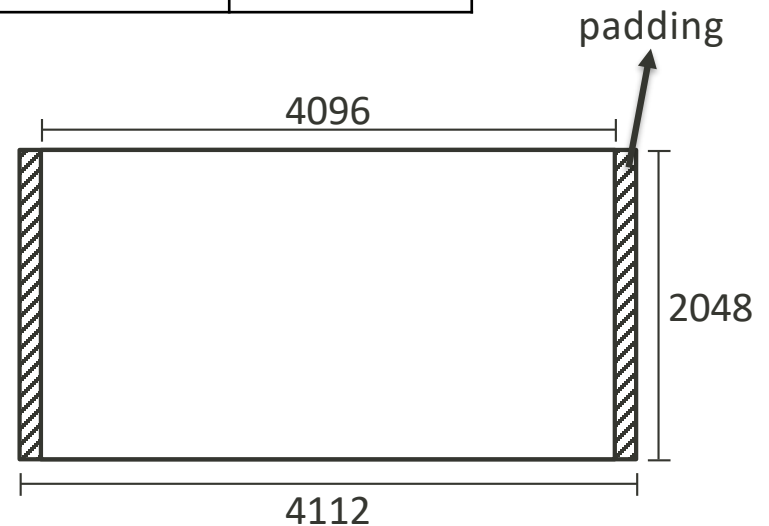
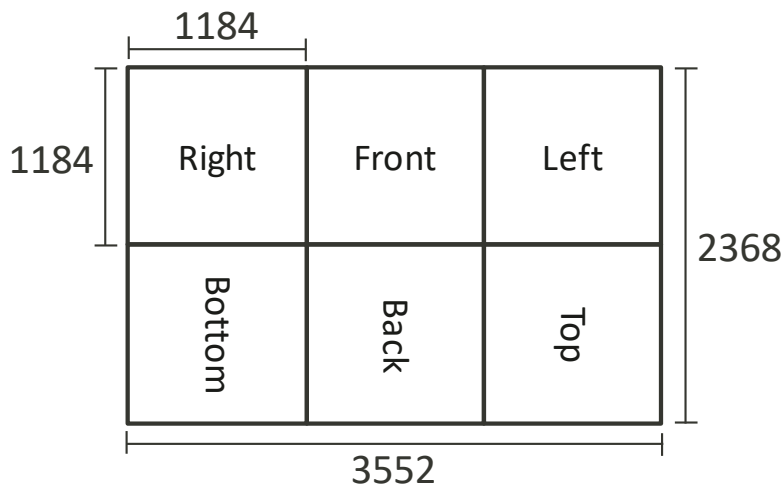


Use geometric neighboring pixels for filtering

Configuration

- Projection setting
 - Compact 3x2 layout without padding
 - The coded samples of the MCP is about 99.88% of the anchor (PERP)

	Face Width	Face Height	Frame Width	Frame Height
Anchor	4096	2048	4112	2048
Proposed	1184	1184	3552	2368



Compression Performance

- Face size: 1184x1184
- Proposed vs. HM anchor

	E2E WS-PSNR				E2E SPSNR-NN			
	Y	U	V	YUV	Y	U	V	YUV
Balboa	-46.6%	-73.2%	-77.0%	-53.7%	-46.6%	-73.2%	-77.0%	-53.7%
Chairlift	-49.3%	-73.1%	-69.3%	-54.8%	-49.3%	-73.1%	-69.3%	-54.7%
KiteFlite	-26.8%	-71.1%	-74.4%	-38.3%	-26.7%	-71.1%	-74.3%	-38.2%
Harbor	-30.3%	-68.7%	-70.2%	-40.1%	-30.0%	-68.8%	-70.2%	-39.9%
Trolley	-24.6%	-61.4%	-67.3%	-34.5%	-24.5%	-61.5%	-67.3%	-34.4%
Average	-35.5%	-69.5%	-71.6%	-44.3%	-35.4%	-69.5%	-71.6%	-44.2%
Encoding time	900.7%							
Decoding time	228.5%							

Compression Performance

- Face size: 1184x1184
- Proposed vs. JEM anchor

	E2E WS-PSNR				E2E SPSNR-NN			
	Y	U	V	YUV	Y	U	V	YUV
Balboa	-17.8%	-61.1%	-63.3%	-28.9%	-17.7%	-61.1%	-63.3%	-28.9%
Chairlift	-27.1%	-51.7%	-49.0%	-32.9%	-27.0%	-51.7%	-49.0%	-32.9%
KiteFlite	-11.3%	-49.2%	-50.1%	-20.9%	-11.2%	-49.1%	-50.0%	-20.8%
Harbor	-12.6%	-46.5%	-49.4%	-21.4%	-12.4%	-46.6%	-49.5%	-21.3%
Trolley	-10.4%	-44.8%	-51.2%	-19.8%	-10.2%	-44.8%	-51.2%	-19.7%
Average	-15.8%	-50.7%	-52.6%	-24.8%	-15.7%	-50.7%	-52.6%	-24.7%
Encoding time	113.0%							
Decoding time	161.5%							

Compression Performance

- Face size: 1280x1280 (multiple of a LCU size)
- Proposed vs. HM anchor

	E2E WS-PSNR				E2E SPSNR-NN			
	Y	U	V	YUV	Y	U	V	YUV
Balboa	-47.0%	-74.1%	-78.2%	-54.3%	-47.0%	-74.1%	-78.2%	-54.3%
Chairlift	-50.1%	-73.7%	-69.6%	-55.5%	-50.0%	-73.7%	-69.6%	-55.4%
KiteFlite	-27.2%	-70.9%	-74.3%	-38.6%	-27.2%	-70.9%	-74.2%	-38.5%
Harbor	-31.3%	-68.2%	-69.5%	-40.7%	-31.0%	-68.2%	-69.5%	-40.5%
Trolley	-25.4%	-60.2%	-66.7%	-34.9%	-25.4%	-60.3%	-66.7%	-34.9%
Average	-36.2%	-69.4%	-71.7%	-44.8%	-36.1%	-69.4%	-71.6%	-44.7%

Compression Performance

- Face size: 1280x1280 (multiple of a LCU size)
- Proposed vs. JEM anchor

	E2E WS-PSNR				E2E SPSNR-NN			
	Y	U	V	YUV	Y	U	V	YUV
Balboa	-18.6%	-62.7%	-65.4%	-29.9%	-18.5%	-62.7%	-65.5%	-29.9%
Chairlift	-28.2%	-54.1%	-50.6%	-34.2%	-28.1%	-54.1%	-50.6%	-34.2%
KiteFlite	-11.9%	-49.6%	-50.8%	-21.5%	-11.8%	-49.4%	-50.7%	-21.4%
Harbor	-13.9%	-46.7%	-49.2%	-22.4%	-13.6%	-46.8%	-49.3%	-22.2%
Trolley	-11.3%	-43.3%	-50.7%	-20.2%	-11.3%	-43.2%	-50.7%	-20.2%
Average	-16.8%	-51.3%	-53.4%	-25.7%	-16.7%	-51.3%	-53.3%	-25.6%

Decoding Time

- Bit-stream decoding time: 25.4x and 3.3x of the HM and JEM anchors
- Projection conversion time: 77.0% of the HM anchors
 - The projection conversion for MCP is implemented based on 360Lib-5.0 with C-level optimizations.
- Total processing (decoding + conversion) time : 2.3x and 1.6x of the HM and JEM anchors

	HM anchors [s]	JEM anchors [s]	Proposed [s]	Proposed vs. HM anchors	Proposed vs. JEM anchors
Bit-stream Decoding	0.24	1.82	6.07	2544.6%	333.9%
Projection Conversion	3.70		2.85	77.1%	
Total Processing	3.95	5.59	9.03	228.5%	161.5%

average processing time per frame

Conclusions

- A MCP format, arranged into a compact 3x2 layout.
- 360° Video specific tools:
 - CU Partition
 - Inter Prediction
 - Intra Prediction
 - Neighboring CU Availability
 - 360°-based In-loop Filters
- With C-level optimizations to reduce the complexity of 360Lib
 - ERP: 50% conversion time reduction
 - Cube: 40% conversion time reduction

- Performance:

Face size	Anchor	E2E WS-PSNR			
		Y	U	V	YUV
1184x1184	HM	-35.5%	-69.5%	-71.6%	-44.3%
	JEM	-15.8%	-50.7%	-52.6%	-24.8%
1280x1280	HM	-36.2%	-69.4%	-71.7%	-44.8%
	JEM	-16.8%	-51.3%	-53.4%	-25.7%



everyday genius